# 3 Time series and dynamically structured problems

*Isn't time just another dimension of the multidimensional smart grid learning problems? How does time and its dependencies interfere with learning? Which properties of supply and demand can be captured from time-series data? How can we synthesise time-series data to make them useful for learning in smart grids?*

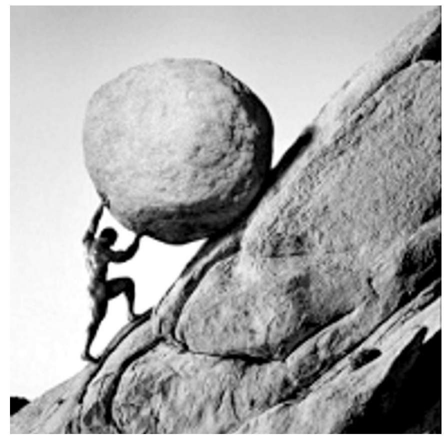

Figure 23: The Sisyphus myth, time and the absurd. Time is the dimension in which the absurd is revealed.

In this chapter, we elaborate on the structural properties of the grid external requests, being those requests driven by meteorological conditions or by loads, generators or storage injections. We focus on how data on such requests can be analysed to understand the grid condition and how data can be synthesised to help predicting future operational grid contexts.

External requests happen in time with different time-dependencies. Although different, time-dependencies share some structural properties that can be expressed mathematically by stochastic processes[13]. Being able to characterize such processes is very important for learning.

Recall that the goal of learning is that of minimizing a loss function, $\mathcal{L}_{\mathcal{D}}(h)$, and that such loss function depends on an unknown distribution $\mathcal{D}$. So far, we have discussed learning methods that relied on minimizing the empirical risk and searched for ERM over a *known* distribution $\mathcal{D}$, either because we have sampled the distribution (we have used training sets for both phase identification and losses estimation) or just because we have assumed $\mathcal{D}$ was known (we have assumed it was Gaussian in state estimation).

Moreover, we have assumed that $\mathcal{D}$ was static in the sense that it did not change with time. This assumption was valid because, in spite of using time-stamped data, time was not a variable in the problems addressed so far; time-series data were considered just as a set of time-stamped data, sampled independently. Note that, in each of the problems addressed so far, we could interchange the chronological order of the data points without any consequence to the results obtained.

Yet, time is not just another variable running in a separate dimension that we can simply ignore. Time imposes a dependency

[13] A stochastic process is defined as a collection of random variables that is indexed by a set, called the index set, that usually is some subset of the real line, such as the natural numbers, giving the index set the interpretation of time.

over variations in all dimensions and conducts such dependencies in one direction only – the arrow of time direction. So, to learn in time, one has to understand time impositions and avoid being trapped in the absurd of ignoring time dependencies when searching for relationships between variables.

Let us illustrate such absurd with the myth of Sisyphus, evoked by the popular image of Fig. 23. Sisyphus rolls an immense boulder up a hill every day and the boulder rolls down the hill every night. It would be absurd to think that he would repeat his job for eternity without learning that it was completely futile. The absurd relies on not recognizing causality in a sequence of events repeated without visible changes: once the boulder is on the top, it always rolls down. Causality needs however to be inferred. Sisyphus does not see the boulder rolling down. However, we would expected him to connect the two events (being on top and rolling down). Such connection is driven by a time-coincidence, impossible to be revealed in a dimension other than time. Time coincidences, or time correlations if we want to use the statistics jargon, are not necessarily caused by existing dependencies – it is repetition (in time) that makes us suspect of causality. The absurd is revealed by repetition, so learning in time is the process of recognizing such absurd (Sisyphus doesn't learn[14]) and interpreting correlation between the two events as a causal dependency.

In this chapter, we will address some of the difficulties imposed by time. Back to the more prosaic contexts of linear regression, time often imposes dependencies between data points that make the error assumptions fail. When data are obtained as time series, the sample often fails the *independence* condition as the random errors are often positively correlated over time. This phenomenon is known as autocorrelation (or serial correlation) and cannot be ignored without consequences to the learning process. Let us elaborate a bit on autocorrelation and its consequences to regression.

A time series is a sequence of measurements made over time for the same variable. Usually the measurements are made at evenly spaced times, for example: daily, hourly, every 15 minutes. Let us consider a variable $x$ measured as a time series of wind speed, with measurements observed periodically. To emphasize that we have measured values over time, we use $t$ as a subscript rather than the previous $k$, i.e., $x_t$ refers to the wind speed $x$ measured in time

[14] I like to think that Sisyphus was punished with the utmost idiocy of not recognizing the absurd by not being cleaver enough to infer causality (he became famous for his cleverness, trickery and for twice cheating death) rather that punished with an eternity of useless efforts that he was conscious about.

period $t$.

Suppose we want to capture the autocorrelation in the wind speed time series. We could model the time series as an autoregressive (AR) process, i.e., as a time-varying random processes where the output variable depends linearly on its own previous values and on a stochastic imperfectly predictable term. The model can be written in the form of a stochastic recurrence equation, as the following:

$$x_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_{t-i} + \varepsilon_t \tag{25}$$

Where $p$ is the order of the autoregression, i.e., the number of immediately preceding values of $x$ that need to be used to predict its value at a given time $t$.

Now suppose we want to regress some grid internal variable $y$ onto the external wind speed variable $x$. Because $x$ is autocorrelated, we expect $y = \beta_0 + \beta x$ to also be autocorrelated, and more importantly the error $\varepsilon$ to be autocorrelated too (not i.i.d., as assumed for ordinary least squares estimates of the beta coefficients). If, to discover $\beta$ in:

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i x_{t-i} + \varepsilon_t \tag{26}$$

we limit $p$, then the model becomes misspecified and we need to assume that the error reflects that limitation by assuming that it is also autocorrelated, i.e.:

$$\varepsilon_t = \rho_0 + \sum_{i=1}^{p} \rho_i \varepsilon_{t-i} + \omega_t. \tag{27}$$

In the first part of this chapter, we will look into possible solutions to regression problems with autoregressive stochastic errors, either by mitigating the effects of autocorrelation on regression or by eliminating its causes, adding independent variables to the regression model.

Yet, before that, we need to address other practical difficulties related to the i.i.d. assumption.

The data in time series do not fail the independence condition only. These data might fail the *identical* condition too, this way violating again the i.i.d. assumption necessary to ordinary least squares solutions. The distribution $\mathcal{D}$ is expected to change over time making the random process a non-stationary stochastic process. See Fig. 24 for a comparison between stationary and non-stationary processes.

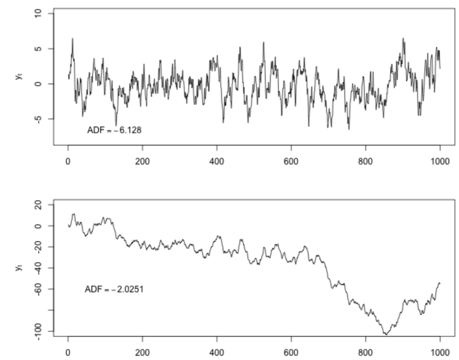


Figure 24: Two time-series processes, one stationary (top plot) and the other non-stationary (bottom plot) [source: en.wikipedia.org/wiki/Stationary_process].

Non-stationarity is a very visible property of many grid external requests. Loads have strong dynamics because human activity has strong dependencies on weather, day-light cycles, energy prices, etc. Generation has also strong dynamics due to thermal and mechanical inertia, besides other strong dependencies on weather, when renewable.

For the learning processes to consider dynamics, such dynamics need to be modelled and translated into a family of distributions $\{\mathcal{D}\}$. Note that a stochastic process is a mathematical object usually defined as a family of random variables $\{X(t)\}, t \in T$. One practical way to model discrete-time stochastic processes (like time series) is to discretize the space value of $X$ and rely on Markov chains to define the probability distribution of $X_t$ over time $t$.

A discrete-time Markov chain is a sequence of random variables $X_t$, $t \in T$ with the Markov property, i.e., a sequence in which the probability of the variable moving from one given state $X_t$ to the next state $X_{t+1}$ is dependent on that previous state alone:

$$\mathbf{Pr}(X_{t+1} = x \mid X_1 = x_1, \ldots, X_t = x_t) = \mathbf{Pr}(X_{t+1} = x \mid X_t = x_t).$$

From the autocorrelation point of view, a Markov chain is an autoregressive process of order $p = 1$. It is, therefore, a stationary process. To model non-stationary processes, we need a Markov chain that is non-homogeneous. A non-homogeneous Markov chain is a chain where,

$$\mathbf{Pr}(X_{t+1} = x \mid X_t = y) \neq \mathbf{Pr}(X_t = x \mid X_{t-1} = y).$$

In Fig. 25 we show an example of a non-homogeneous Markov chain where the space value is discretized into 3 states and the horizon $T$ is discretized into 96 periods of 15 min to represent the intra-day dynamics of a load.

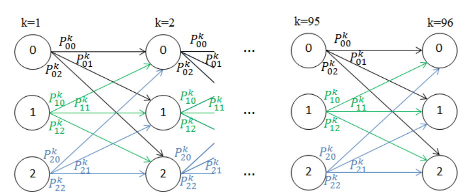


Figure 25: Illustration of a non-homogeneous Markov chain with 3 states and 96 time periods.

Now suppose we want to estimate the state of a grid whose loads are represented by non-homogeneous Markov chains. In the state estimation example of the previous chapter, latency in refreshing load measurements was responsible for significant errors in pseudo-measurements, which we assumed to be Gaussian with high variance, remember? How could we now model latency in pseudo-load measurements? (suppose we have real-time measurements in time $t$ and pseudo-measurements in $t-1$). The pseudo-measurement errors will now have both a mean and a variance term (not just variance) and the mean error term will be dependent of the grid's previous state, as estimated.

In the second part of this chapter, we will look into possible solutions to state estimation where loads are modelled as non-stationary stochastic processes based on time-series data.

## 3.1 Regression over quasi-stationary processes – AR models

Suppose we have times-series data both on a given electrical quantity $y$ and on a set of stochastic grid variables $X$, and want to project $y$ onto $X$ to predict $y$ based on forecasts of $X$.

The task of projecting available forecasts into other grid variables is an important one. Forecasts are usually obtained for variables of significant dynamics and exogenous dependencies – such as large loads, solar and wind power generation output – and we need to anticipate the consequences of such forecasts to the grid operation by projecting them into relevant electrical quantities such as nodal voltages and branch currents – the variables from which the grid condition is usually assessed.

Assuming that the stochastic process $\{X(t)\}$ ruling such forecast is one that can be approximated by an autocorrelated process,

$$X_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \varepsilon_t$$

a regression equation could be expressed as in (26) by:

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i X_{t-i} + \varepsilon_t.$$

Because $X_t$ is autocorrelated, we expect $y_t$ to be autocorrelated as well. But more importantly is that we should expect the errors $\varepsilon_t$ to be autocorrelated too.

The reason why the error might be autocorrelated is that the model relating $y$ and $X$ measurements might be misspecified, either because the order $p$ of the model is limited, or just because $X$ does not include all relevant independent variables. The misspecified relationships between $y$ and $X$ are absorbed into the error term making it autocorrelated as well.

The autocorrelation of errors is a violation of the assumption of independence, and therefore creates a theoretical (and practical) difficulty to the ordinary least-squares estimates of the beta coefficients. When the assumption of the Gauss-Markov theorem does not hold, the question seems to be that of being able to accurately predict the grid condition based on times series data of dynamic variables.

The problem can be stated as follows:

**Consider a set of stochastic variables $X$ and an electrical quantity $y$ dependent on such variables for which time-series data are available (both for the variables and the quantity) and discover an accurate predictor for $y$ based on the forecasts of $X$.**

We may approach the problem starting from an assumption that simplifies its solution. We may assume that the process generating the residuals $\varepsilon_t$ is a stationary first-order autoregressive process, i.e., a process with a structure like:

$$\varepsilon_t = \rho\varepsilon_{t-1} + \omega_t,\ |\rho| < 1, \tag{28}$$

with $\omega_t$ being white noise.

If the errors are first-order autoregressive, one may rewrite (28) to isolate the white noise as $\varepsilon_t - \rho\varepsilon_{t-1} = \omega_t$ and transform the model of (26) into the quasi-difference model that depends on white noise only, i.e.:

$$y_t - \rho y_{t-1} = \beta_0(1-\rho) + \beta(X_t - \rho X_{t-1}) + \omega_t. \tag{29}$$

This transformation is expected to generate residuals that are not autocorrelated or are hopefully less correlated than before the transformation. This transformation is known as the Cochrane–Orcutt procedure, and consists in regressing over the transformed variables

$$\begin{aligned} y_t^* &= y_t - \rho y_{t-1} \\ X_t^* &= X_t - \rho X_{t-1} \end{aligned}$$

with white noise $\omega_t$ to determine $\beta_0^* = \beta_0(1-\rho)$ and $\beta^* = \beta$ in

$$y_t^* = \beta_0^* + \beta^* X_t^* + \omega_t.$$

For the transformation to be carried out, we first need to estimate $\rho$, which we may do by autoregressing over the residuals $\varepsilon_t$.

In the following example we show how to mitigate the effect of autocorrelation in estimating grid conditions when such conditions depend on dynamic variables, and what the expected consequences of such mitigation might be. We will then discuss how to reduce estimation errors by adding new independent stochastic variables to the model – removing the cause of autocorrelation instead of mitigating the effects of it – and how that might lead us from regression to dynamic state estimation – the topic of the second part of this chapter.

**Regression over Autoregressive Forecasts Example**: Let us use the five-bus system of the "kite" network and assume we have access to real time measurements of both the active power injection in bus 1 (from where a large wind farm is connected), $P_1^{inj}$, and the magnitude (and sign) of the current flowing in line 1-2, $I_{12}$. Assuming we have access to wind power forecasts for the next 12 hours, we want to anticipate the hourly current flows in line 1-2 as accurately as possible. Fig. 26 shows the network model with the wind farm connection and the grid measurement locations.

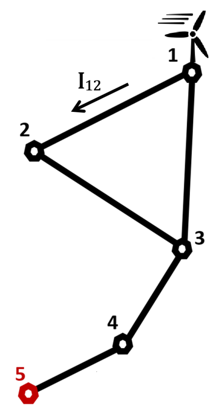


Figure 26: Regressing $I_{12}$ onto wind farm AR injection from bus 1.

The obvious way to project the injection forecast into the current flow is to use times-series measurements on both such variables and regress the first over the second variable based on such measurements.

We could express $y_t \doteq I_{12}(t)$ onto $x_t \doteq P_1^{inj}(t)$ and obtain $\beta$ and $\beta_0$ with least squares as follows:

$$\begin{pmatrix} \beta_0 \\ \beta \end{pmatrix} = (X_t^T X_t)^{-1} X_t^T y_t \tag{30}$$

where,

$$X_t = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_T \end{pmatrix}.$$

The least squares solution obtained with the hourly time-series data of Fig. 27 is shown in Fig. 28. The figure shows graphically the data on both the bus power injection, $P_1^{inj}$, and the corresponding current magnitude in the bus outgoing line, $I_{12}$. The figure also shows the output of the linear regression and the corresponding residuals when plotted against time.

The non-random trend of the residuals in time is suggestive of significant autocorrelation in the errors − a consequence of misrepresented dynamics in the model and a violation of the i.i.d. assumption. The suggestion, derived from the subjective analysis of the trend, can be supported by statistical evidence. The Durbin-Watson statistic provides such support by approximating with $DW$ the value of $2 - 2\rho$, where $DW$ is defined as:

$$DW = \frac{\sum_{t=2}^{T} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{T} \varepsilon_t^2} = 0.229 \rightarrow \rho \approx 0.885$$

```
[x y] =

  -0.2128    0.1202
   0.1282    0.1959
   0.1106    0.1801
   0.6675    0.2547
   0.4977    0.1508
   0.1015    0.0291
  -0.0650   -0.0335
  -0.2546   -0.0304
  -0.3703    0.0427
  -1.2655   -0.1369
  -1.4417   -0.1597
  -1.2293   -0.0772
  -0.1998    0.2258
  -0.4622    0.1098
   0.1074    0.2412
   1.0280    0.5023
   0.6125    0.3312
   0.9286    0.5136
   1.7137    0.7404
   1.2350    0.6556
   1.3946    0.7686
   0.6856    0.5978
   0.9820    0.6858
   1.0412    0.8254
```

Figure 27: Vectors $x_t$ and $y_t$ used in the example as $P_1^{inj}$ and $I_{12}$, respectively.

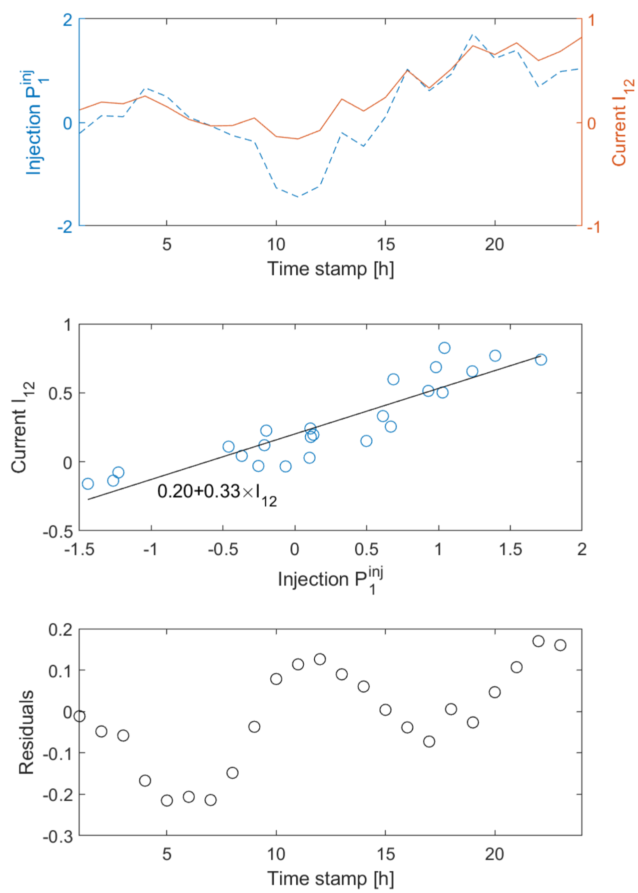


Figure 28: Active power injection in the wind farm connection bus, $P_1^{inj}$, and corresponding current flow magnitude (with sign) in the outgoing line, $I_{12}$, versus time (top plot); the injected power versus the current flow with the ordinary least squares regression line overlaid (middle plot); and a scatter plot of the residuals versus time (bottom plot).

The high autocorrelation coefficient $\rho$ found in the residuals of the example is a consequence of the built-in wind speed dynamics, created for this example as lag-1 autoregressive, i.e., the time-series data generating process was AR(1). But not just that. It is also a consequence of the existing dynamics of the load demanded from the other grid buses, which we did not represent in the regression model.

The model obtained with (30) does not explicitly addresses some important dynamic variables, such as the loads (it ignores them) and, therefore, cannot be expected to perform very well in estimating the future current flows despite the wind forecast being accurate and the residuals found by the linear regression being small.

Fig. 29 shows – as a black dashed line – the results of the 12h-ahead projection for the current flow $I_{12}$ obtained with (30) for 12h-ahead accurate wind forecasts. Not very good, right?

The effect of autocorrelation in residuals can however be mitigated. A usual way of doing it is to take the Cochrane–Orcutt approach. Fig. 29 also compares the results obtained with (30) with the (much better) results obtained after three iterations of the Cochrane–Orcutt procedure.

The following summarises the main steps of the Cochrane–Orcutt transformation procedure, as implemented:

**Regress** $y_t$ over $x_t$ and store the residuals $\varepsilon_t$.

**Assign** $y_t^* := y_t$ and $x_t^* := x_t$.
**while** *autocorrelation is present* **do**

> Estimate $\rho$ with regression through the origin for $\varepsilon_t$.
>
> Transform $y_t^* = y_t - \rho y_{t-1}$, $x_t^* = x_t - \rho x_{t-1}$.
>
> Regress $y_t^*$ over $x_t^*$ and estimate $\beta_0^*$ and $\beta^*$.
>
> Transform $\beta_0 = \beta_0^*/(1-\rho)$ and $\beta = \beta^*$.
>
> Update the residuals $\varepsilon_t = y_t - (\beta_0 + \beta x_t)$.

**end**

The Cochrane–Orcutt procedure resulted in a good solution to mitigate the effects of autocorrelation and turned out to provide acceptable projections for the current flow profile in this example. That can be confirmed in Fig. 29 by noticing that the red-dashed line in much closer to the red line than the black-dashed one.

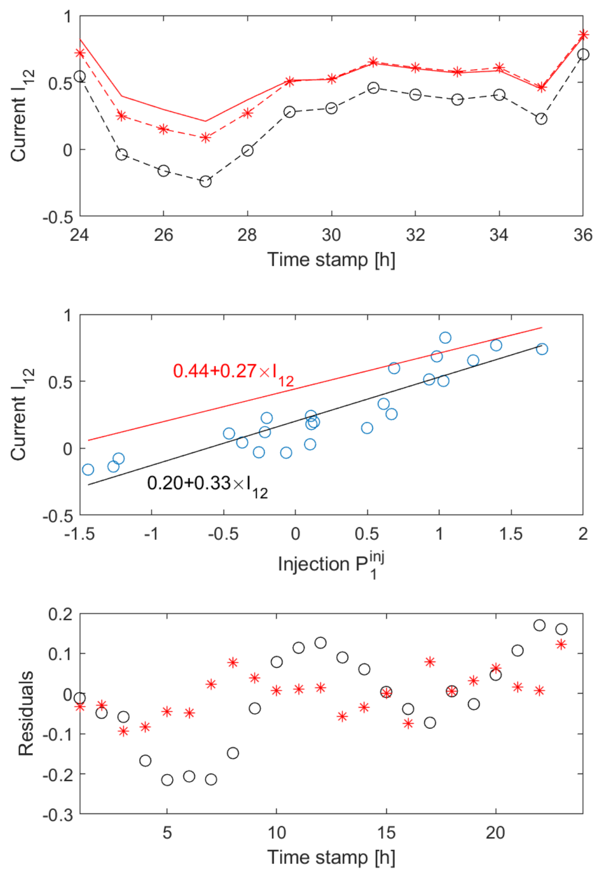


Figure 29: 12h-ahead projection for current flow $I_{12}$ (top plot). The red solid line is the *a posteriori* current measurement and the dashed lines are the *a priori* values obtained by regression over the forecast of $P_1^{\widehat{inj}}$: with ordinary least squares (in black) and after the Cochrane–Orcutt transformation (in red). The middle and bottom plots show the corresponding (black/red) regression equations, and the obtained residuals versus time.

Instead of mitigating the effects of autocorrelation, we could try to avoid its causes – and the causes might simply be that some important dynamic effect has been neglected by the regression model.

When autocorrelation is detected in the residuals from a model, it suggests that the model is misspecified (i.e., in some sense wrong). One important form of misspecification occurs when the true model is dynamic and the used model wrongly assumes that it is not – that it is static instead. Experience supports some evidence that when Durbin-Watson statistic $DW$ is significantly different from 2, one should look for a dynamic formulation of the problem instead of reverting to a transformation of the Cochrane-Orcutt type.

In our example, we formulated the regression problem as a static problem, explicitly ignoring its autoregressive nature. If we recall (25) and formulate the problem as AR(1), adding $y_{t-1}$ to the regression model, as follows:

$$y_t = \beta_0 + \beta_1 x_t + \alpha y_{t-1} + \varepsilon_t \tag{31}$$

we could then express $I_{12}(t)$ on both $I_{12}(t-1)$ and $P_1^{inj}(t)$ and use least squares to parameterise the model as follows:

$$\begin{pmatrix}\beta_0\\ \beta_1\\ \alpha\end{pmatrix} = (X_t^T X_t)^{-1} X_t^T y_t \tag{32}$$

Where,

$$X_t = \begin{pmatrix} 1 & x_2 & y_1 \\ 1 & x_3 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_T & y_{T-1} \end{pmatrix}. \tag{33}$$

By doing so, we would slightly improve the results of the $I_{12}$ projection when compared to the original approach, but lose the ability to project $I_{12}$ more than one period ahead of time. Note that $I_{12}(T+k)$ needs to be estimated with the value of $I_{12}(T+k-1)$. Even though, we still could not deliver an acceptable solution[15]. The reason is that by doing as proposed we kept the loads out of the model, ignoring their important dynamics.

In Fig. 31, we compare results from the static ordinary least squares (OLS) approach of (30) with two AR(1) approaches: one

[15]We could further improve the forecast accuracy by also adding to the model $P_1^{inj}(t-1)$ these way allowing to express the differences in consecutive outputs on the differences in consecutive inputs

| $P_1(t)$ | $I_{12}(t-1)$ | $\Sigma P_i(t)$ i=1…4 |
|---|---|---|
| 0.1282 | 0.1202 | -1.3268 |
| 0.1106 | 0.1391 | -1.9596 |
| 0.6675 | 0.1853 | -1.8262 |
| 0.4977 | 0.3585 | -2.3787 |
| 0.1015 | 0.3487 | -3.2403 |
| -0.0650 | 0.2894 | -5.0096 |
| -0.2546 | 0.3812 | -5.1187 |
| -0.3703 | 0.3280 | -5.2445 |
| -1.2655 | 0.3007 | -6.4463 |
| -1.4417 | 0.1088 | -5.5240 |
| -1.2293 | -0.0522 | -4.6474 |
| -0.1998 | -0.0503 | -3.2216 |
| -0.4622 | 0.1919 | -3.6595 |
| 0.1074 | 0.1422 | -1.8665 |
| 1.0280 | 0.1823 | -0.7789 |
| 0.6125 | 0.3983 | -1.7457 |
| 0.9286 | 0.3407 | 0.0111 |
| 1.7137 | 0.3006 | 2.6552 |
| 1.2350 | 0.3431 | 1.4732 |
| 1.3946 | 0.2820 | 1.2509 |
| 0.6856 | 0.3540 | 0.5569 |
| 0.9820 | 0.1758 | 2.3538 |
| 1.0412 | 0.1257 | 3.4554 |
| -0.7289 | 0.0544 | 1.8081 |
| -1.0981 | -0.3988 | 1.6670 |
| -1.3395 | -0.5099 | 2.0330 |
| -0.6390 | -0.6204 | 2.3591 |
| 0.2368 | -0.4137 | 2.8667 |
| 0.3183 | -0.1636 | 2.8142 |
| 0.7811 | -0.1318 | 2.7981 |
| 0.6267 | 0.0240 | 3.2758 |
| 0.5151 | -0.0669 | 4.8990 |
| 0.6228 | -0.2385 | 3.4075 |
| 0.0797 | -0.0787 | 1.5617 |
| 1.5380 | -0.1065 | 4.3663 |

Figure 30: Data used in the example of Fig. 31. The regression models were fitted with the time-series data from the first 23 rows (in black) and were tested for forecast accuracy with following 12 rows of data (in rede).

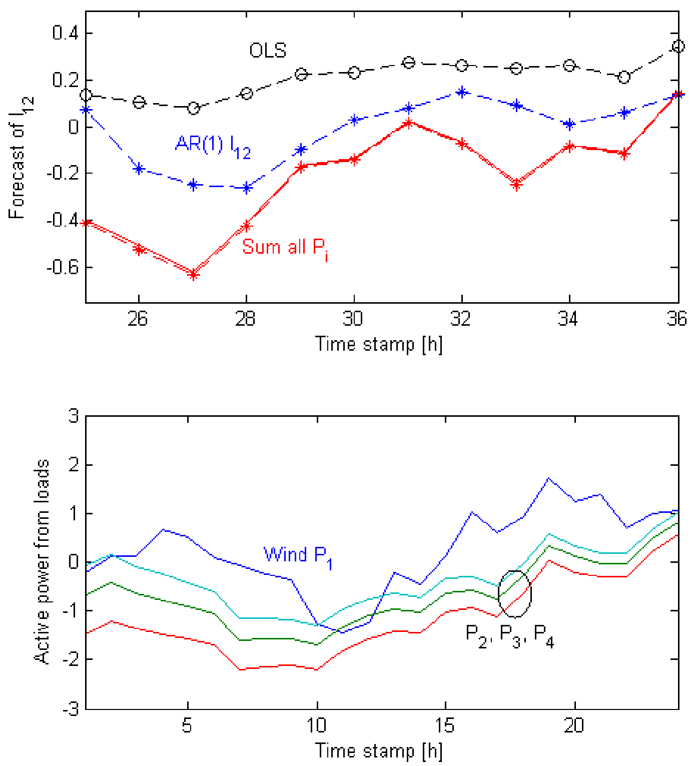


Figure 31: 12h-ahead line current projections (top plot). The red solid line is the *a posteriori* current measurement and the dashed lines are the *a priori* values obtained by regression (1) over the forecast of $P_1^{inj}$ with OLS – in back; (2) over the forecast of $P_1^{\hat{inj}}$ and the lag-1 current flow values, $I_{12}(t-1)$ – in blue; and (3) over $P_1^{\hat{inj}}$ and the sum of all bus injections – in red. The bottom plot shows the load and wind power injections $P_{1-4}^{inj}$ versus time, used to train the model.

that just regresses on previous current values, $I_{12}(t-1)$, as proposed in (32), and another that replaces $I_{12}(t-1)$ in the matrix $X_t$ in (33) by a new column with the sum of all bus injections versus time (including the wind injection), $\sum_{i=1,..4} P_i$.

The plot of Fig. 31 was obtained by training the model with data depicted in Fig. 30 and by applying the trained model to forecasts generated to be significantly different from the training data – the idea was to find a situation in which the projection errors would be high. Note that errors are now much higher than before. The new data and forecasts are depicted in the figure's bottom plot.

The figure illustrates the importance of adding an independent load variable to the model by making obvious that such addition reduces the projection error to negligible values (the posteriori and the forecast with Load Sum results are very similar).

Because the dynamics of the three different loads are very similar to each other (in this example, see the bottom plot), any combination of load variables would perform equally well in reducing projection errors – we opted by the sum of all loads because in practice it would be easier to forecast, and easier to measure, as a simple conjugate function of the outgoing feeder current from bus 5.

In practice, load dynamics of the same type of load-use tend to be similar but may be very different from the dynamics of other load-uses. Domestic loads have similar dynamics which are very different from commercial or industrial ones. In our example, if load dynamics were substantially different from each other, we would have to measure and forecast each one of them in order to project the current flow with acceptable accuracy[16]. That could be possible, although possibly not very wise. Reasons are that:

- The loads' are usually not AR, might not be stationary, and even if we assume they were, they would be high order AR in order to reproduce existing daily/weekly periodicity in their use – and that would make the regression task quite a pain;
- By adding all load variables as independent variables to the model, the model would become so tied-up that regression would lose practical interest – with advantage to the grid operator, we could perhaps address such problem as a dynamic state-estimation problem, since we would probably have enough data to create some kind of observability.

[16]Instead of forecasting them all, we could use PCA to reduce the number of necessary independent load forecast and still explain most of the variance in the current flow.

In the following, we will elaborate on how to model load dynamics in order to use such models in state estimation.